REMARKS

In the final Office Action, the Examiner rejects claims 1-3, 5, 7-14, and 16-21 under 35 U.S.C. § 103(a) as unpatentable over NAJORK et al. (U.S. Patent No. 6,321,265; referred to hereinafter as "NAJORK '265") in view of HOFFERT et al. (U.S. Patent No. 6,374,260); rejects claims 22-24 under 35 U.S.C. § 103(a) as unpatentable over NAJORK '265 in view of NAJORK et al. (U.S. Patent No. 6,351,755; referred to hereinafter as "NAJORK '755"); and rejects claim 26 under 35 U.S.C. § 103(a) as unpatentable over NAJORK '265. Applicants respectfully traverse these rejections.[1] Claims 1-3, 5, 7-14, 16-24, and 26 remain pending.

INTERVIEW SUMMARY

In accordance with Applicants' duty to provide information regarding the substance of an interview, an interview was held between Applicants' representative and Examiner Bashore on June 20, 2007. Applicants would like to thank Examiner Bashore for the courtesies extended during the interview. During that interview, the rejection based on NAJORK '265 in view of HOFFERT et al. was discussed. In particular, Applicants' representative discussed that the disclosure of NAJORK '265 specifically discloses storing links in queues by host (where each queue is associated with a host) and selecting a link to crawl from the queue having a minimal next download time (see, for example, Figs. 3 and 4 of NAJORK '265). However, Applicants' claims, such as claim 1,

[1] As Applicants' remarks with respect to the Examiner's rejections are sufficient to overcome these rejections, Applicants' silence as to assertions by the Examiner in the Office Action or certain requirements that may be applicable to such rejections (e.g., whether a reference constitutes prior art, motivation to combine references, assertions as to dependent claims, etc.) is not a concession by Applicants that such assertions are accurate or such requirements have been met, and Applicants reserve the right to analyze and dispute such assertions/requirements in the future.

recites more than merely storing links by host. Applicants' claim 1, for example, recites, *inter alia*, grouping the plurality of links to hyperlinked documents by host; grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host; sorting the hosts in each bucket based on a stall time of each host; selecting a host from one of the buckets to crawl next according to the stall time of the host; and crawling a hyperlinked document from the selected host. NAJORK '265 in no way discloses or suggest this combination of features.

While no agreement was reached with respect to allowable subject matter, the Examiner agreed to reconsider the outstanding rejections upon the filing of this Request for Reconsideration.

REJECTION UNDER 35 U.S.C. § 103(A) BASED ON NAJORK '265 AND HOFFERT ET AL.

Claims 1-3, 5, 7-14, and 16-21 stand rejected under 35 U.S.C. § 103(a) as allegedly unpatentable over NAJORK '265 in view of HOFFERT et al. Applicants respectfully traverse this rejection.

Independent claim 1 is directed to a computer implemented method of crawling hyperlinked documents. The method includes sending a request for additional links to hyperlinked documents to a link manager; receiving a plurality of links to hyperlinked documents to be crawled, where the plurality of links is selected by the link manager based on priority; grouping the plurality of links to hyperlinked documents by host; grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host; sorting the hosts in each bucket based on a stall time of each host; selecting a host from one of the buckets to crawl next according to the stall time of the

host; and crawling a hyperlinked document from the selected host. NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination, do not disclose or suggest this combination of features.

For example, NAJORK '265 and HOFFERT et al. do not disclose or suggest grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host. The final Office Action does not address this feature. Accordingly, a *prima facie* case of obviousness has not been established with respect to claim 1. Applicants' representative discussed this fact with the Examiner in the Interview on June 20, 2007.

Applicants note that the Office Action, dated May 12, 2005, admits that NAJORK '265 does not disclose grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host (see p. 6). The current Office Action relies on HOFFERT et al. for allegedly disclosing that the plurality of links (pages) to be crawled is selected based on priority (Office Action, p. 3). Regardless of the veracity of this allegation, HOFFERT et al. does not disclose or suggest grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host, as recited in claim 1.

With respect to the above arguments, the final Office Action alleges:

> Najork '265 teaches a Frontier data structure providing links to a web crawler, as well as teaching receiving a plurality of links to hyperlinked documents to be crawled. Najork '265 teaches grouping the plurality of links to hyperlinked documents by host. Since "buckets" serve to group items accordingly, grouping links by host can be reasonably interpreted as grouping into buckets. Najork '265 also teaches selecting a host to crawl next according to a stall time of the host

(final Office Action, p. 12). Applicants respectfully disagree with these allegations.

It appears that the final Office Action indicates that NAJORK '265's disclosure of grouping links by host corresponds to both grouping the plurality of links to hyperlinked documents by host and grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host, as recited in claim 1. Applicants submit that this position is unreasonable and not supported by the NAJORK '265 disclosure.

Applicants respectfully request that the Office address the above feature of claim 1 or withdraw the rejection.

Since NAJORK '265 and HOFFERT et al. do not disclose or suggest grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host, NAJORK '265 and HOFFERT et al. cannot disclose or suggest sorting the hosts in each bucket based on a stall time of each host, as also recited in claim 1. The final Office Action does not address this feature. Accordingly, a *prima facie* case of obviousness has not been established with respect to claim 1.

Applicants note that the Office Action, dated May 12, 2005, admits that NAJORK '265 does not disclose sorting the hosts in each bucket based on a stall time of each host (see p. 7). The Office Action relies on HOFFERT et al. for allegedly disclosing that the plurality of links (pages) to be crawled is selected based on priority (Office Action, p. 3). Regardless of the veracity of this allegation, HOFFERT et al. does not disclose or suggest sorting the hosts in each bucket based on a stall time of each host, as recited in claim 1.

Applicants respectfully request that the Office address the above feature of claim 1 or withdraw the rejection.

For at least the foregoing reasons, Applicants submit that claim 1 is patentable over NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination.

Claims 2, 3, 5, and 7-9 depend from claim 1. Therefore, these claims are patentable over NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination, for at least the reasons given above with respect to claim 1.

Independent claims 10, 12, and 20 recite features similar to (yet possibly of different scope than) features identified above with respect to claim 1. Therefore, these claims are patentable over NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination, for at least reasons similar to reasons given above with respect to claim 1.

Claim 11 depends from claim 10. Therefore, this claim is patentable over NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination, for at least the reasons given above with respect to claim 10.

Claims 13, 14, and 16-19 depend from claim 12. Therefore, these claims are patentable over NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination, for at least the reasons given above with respect to claim 12.

Claim 21 depends from claim 20. Therefore, this claim is patentable over NAJORK '265 and HOFFERT et al., whether taken alone or in any reasonable combination, for at least the reasons given above with respect to claim 20.

REJECTION UNDER 35 U.S.C. § 103(A) BASED ON
NAJORK '265 AND NAJORK '755

Claims 22-24 stand rejected under 35 U.S.C. § 103(a) as allegedly unpatentable over NAJORK '265 in view of NAJORK '755. Applicants respectfully traverse this rejection.

Independent claim 22 is directed to a computer implemented method of crawling hyperlinked documents. The method includes storing a plurality of links to hyperlinked documents to be crawled; determining that more links to hyperlinked documents are desired; sending requests to multiple link managers for more links to hyperlinked documents; receiving additional links to hyperlinked documents from the link managers; selecting a host to crawl next according to a stall time of the host; and crawling a hyperlinked document from the selected host. NAJORK '265 and NAJORK '755, whether taken alone or in any reasonable combination, do not disclose or suggest this combination of features.

For example, NAJORK '265 and NAJORK '755 do not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents. The final Office Action admits that NAJORK '265 does not disclose this feature and relies on col. 16, lines 10-26, of NAJORK '755 for allegedly disclosing "queues containing URL's from multiple server hosts, as well as teaching multiple queues implemented in various scenarios" (final Office Action, p. 9). Applicants submit that this allegation (regardless of its veracity) does not address the above feature of claim 22.

That is, claim 22 does not recite queues containing URLs from multiple server hosts or multiple queues being implemented in various scenarios. Instead, claim 22 specifically recites, *inter alia*, sending requests to multiple link managers for more links

to hyperlinked documents. Thus, the final Office Action's allegation with respect to NAJORK '755, regardless of its veracity, does not address the above feature of claim 22. Thus, a *prima facie* case of obviousness has not been established with respect to claim 22.

Nevertheless, at col. 16, lines 10-26, NAJORK '755 discloses:

> In the second exemplary embodiment described above, when crawling in a network with a relatively small number of host computers, such as in an Intranet, some queues may be empty while other queues may contain URL's for multiple server hosts. Thus, in the second embodiment, parallelism may not be efficiently maintained, since the threads associated with the empty queues will be idle. The third embodiment described makes better use of thread capacity, on average, by dynamically reassigning queues to whichever hosts have pages that need processing. In both of these exemplary embodiments the same politeness policies may be enforced, whereby the web crawler not only does not submit overlapping download requests to any host, it waits between document downloads from each host for a period of time. The wait time between downloads from a particular host may be a constant value, or may be proportional to the download time of one or more previous documents downloaded from the host.

This section of NAJORK '755 discloses dynamically reassigning queues to hosts that have pages that need processing. This section of NAJORK '755 does not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents, as recited in claim 22.

While the final Office Action admits that the NAJORK '265 does not disclose the above feature of claim 22, the final Office Action also relies on Figs. 2-4 and col. 5, line 53 to col. 6, line 6, of NAJORK '265 for allegedly disclosing this feature (final Office Action, p. 12). Applicants respectfully disagree with this interpretation of NAJORK '265.

In Fig. 2, NAJORK '265 depicts a group of first-in, first-out queues 128-0 to 128-n formed between a demultiplexer 126 and a multiplexer 124. This figure of NAJORK '265 does not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents, as recited in claim 22. In fact, this section of NAJORK '265 does not even relate to sending requests.

In Fig. 3, NAJORK '265 depicts an ordered set data structure 134 for keeping track of the queues that are waiting to be serviced by threads (col. 6, lines 19-21). This figure of NAJORK '265 does not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents, as recited in claim 22. In fact, this section of NAJORK '265 does not even relate to sending requests.

In Fig. 4, NAJORK '265 depicts a flow chart for enqueuing URLs into a set of n queues using a set of k threads (col. 6, lines 57-60). This figure of NAJORK '265 does not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents, as recited in claim 22. In fact, this section of NAJORK '265 does not even relate to sending requests.

At col. 5, line 53 to col. 6, line 6, NAJORK '265 discloses:

> Given a set of URL's, the web crawler 102 begins downloading documents by enqueuing the URL's into appropriate queues 128. Multiple threads 130 are used to dequeue URL's out of the queues 128, to download the corresponding documents or web pages from the world wide web and to extract any new URL's from the downloaded documents. Any new URL's are enqueued into the queues 128. This process repeats indefinitely or until a predetennined stop condition occurs, such as when all URL's in the queues have been processed and thus all the queues are empty. Multiple threads 130 are used to simultaneously enqueue and dequeue URL's from multiple queues 128. During the described process, the operating system 120 executes an Internet access procedure 122 to access the Internet through the communications interface 104.

> The web crawler's threads substantially concurrently process the URL's in the queues. When the web crawler is implemented on a multiprocessor, some of the threads may run concurrently with each other, while others run substantially concurrently through the services of the multitasking operating system 120.

This section of NAJORK '265 discloses downloading documents or web pages and extracting new URLs from the downloaded documents or web pages. This section of NAJORK '265 does not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents, as recited in claim 22

For at least the foregoing reasons, Applicants submit that claim 22 is patentable over NAJORK '265 and NAJORK '755, whether taken alone or in any reasonable combination.

Independent claim 23 recites features similar to (yet possibly of different scope than) features recited above with respect to claim 22. Therefore, Applicants submit that claim 23 is patentable over NAJORK '265 and NAJORK '755, whether taken alone or in any reasonable combination, for at least reasons similar to the reasons given above with respect to claim 22.

Claim 24 depends from claim 23. Therefore, Applicants submit that this claim is patentable over NAJORK '265 and NAJORK '755, whether taken alone or in any reasonable combination, for at least the reasons given above with respect to claim 23.

REJECTION UNDER 35 U.S.C. § 103(A)
BASED ON NAJORK '265

Claim 26 stands rejected under 35 U.S.C. § 103(a) as allegedly unpatentable over NAJORK '265. Applicants respectfully traverse this rejection.

Independent claim 26 is directed to a method for crawling hyperlinked documents. The method includes grouping links to hyperlinked documents by host, where each host is associated with a stall time; grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host; sorting the hosts in each bucket based on the stall time of each host; and identifying a host to crawl by examining the buckets in descending order based on the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than a current time. NAJORK '265 does not disclose or suggest this combination of features.

For example, NAJORK '265 does not disclose or suggest grouping hosts into buckets according to a number of hyperlinked documents to be crawled at each host. The final Office Action does not address this feature. Accordingly, a *prima facie* case of obviousness has not been established with respect to claim 26. Applicants' representative discussed this fact with the Examiner in the Interview on June 20, 2007.

NAJORK '265 does not further disclose or suggest sorting the hosts in each bucket based on the stall time of each host, as also recited in claim 26. The Office Action does not address this feature. Accordingly, a *prima facie* case of obviousness has not been established with respect to claim 26.

Applicants respectfully request that the Office address the above features of claim 26 or withdraw the rejection.

For at least the foregoing reasons, Applicants submit that claim 26 is patentable over NAJORK '265.

In view of the foregoing remarks, Applicants respectfully request the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

By: /John E. Harrity/
John E. Harrity
Registration No. 43,367

Date: July 3, 2007

11350 Random Hills Road
Suite 600
Fairfax, Virginia 22030
(571) 432-0800

Customer Number: 44989